

Building up a Common Data Infrastructure[©]

Mark R. Sinclair

VERIDIAN
Information Solutions Division
2101 Executive Drive; Tower Box 51
Hampton, Virginia 23666
U.S.A.
+1 (757) 825-4097

mark.sinclair@veridian.com

Dr. Andreas Tolk*

Virginia Modeling, Analysis & Simulation Center
Old Dominion University
7000 College Drive
Suffolk, Virginia 23435
U.S.A.
+1 (757) 686-6203

atolk@odu.edu

ABSTRACT

While the value of data for an individual study effort is well understood by the analytic community at large, aggregated worth of data is still astonishingly undervalued by many members of the OR study community. Data can be described as the fundamental elements of information and knowledge that comprises the corporate whole – consequently its aggregated value particularly when addressed in a context larger than an individual study is significantly greater than the sum of the parts.

Obtaining data is indispensable. To be effective it must be a continuous process within every study and can be not only very time consuming but also a very expensive factor in the total cost of a study effort. With the aggregate of available data growing with every study the situation becomes even more complex and the case for agreed community wide data management standards and techniques is made even stronger. Without these standards the analyst's ability to find the necessary data for an individual study effort by traditional means decreases exponentially and the ability to reuse existing data in future studies is reduced thereby increasing the cost of data.

To help the analyst to face these challenges, the NATO Code of Best Practice for Assessment of Command and Control (COBP) introduced a Data Section. This section already defines the application domains of data engineering, meta data modelling and efficient data re-use. However, the deeper value of these additional efforts – albeit a burden for the single study, especially for the initial efforts at introducing the respective techniques and tools – clearly show up when being seen in the broader context of multiple studies dealing with related topics.

This paper extends the application of the COBP data section beyond the scope of a single study into the broadened study community domain, including other Operational Analysts, C3I System Developers, Social Scientists, etc. Therefore, in this paper the necessary methodologies for applying the ideas of the COBP data section, thus enabling the reuse of data across different studies, will be highlighted. A case will be made for a user community requirement for a common data infrastructure including some first ideas for technical implementations.

Key Words: *Data Engineering, Data Mining, Data Farming, Data Re-Use, Meta Data Modelling, Information Repository, Information Resource Dictionary System (IRDS).*

* The contributions to this paper have been conducted on behalf of the Industrieanlagenbetriebsgesellschaft mbH (IABG), Einsteinstr. 20, 85521 Ottobrunn, Germany, where Dr. A. Tolk worked until March 2002.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 00 DEC 2003		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Building up a Common Data Infrastructure				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) VERIDIAN Information Solutions Division 2101 Executive Drive; Tower Box 51 Hampton, Virginia 23666 U.S.A.; Virginia Modeling, Analysis & Simulation Center Old Dominion University 7000 College Drive Suffolk, Virginia 23435 U.S.A.				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM001657., The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 27	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

1.0 INTRODUCTION

The role of data and its importance is acknowledged as fundamental to the conduct of a successful and intellectually sound study. However, in practice data often is neglected during the study preparations. Data is often seen only as something necessary to feed the respective tools and models to be used in the study. It is interesting that the tools and models are usually seen to be of high value whereas the data just is something that is needed “in addition” – not as the fuel that makes the tools run. It is of no great surprise that this view was represented in the first version of the *NATO Code of Best Practice (COBP) for Command and Control Assessment*. Although it is very clearly stated that tools are only as good as the data – and therefore beside the processes of verification, validation, and accreditation (VV&A) for tools, a processes of verification, validation, and certification (VV&C) for data are needed – the requirements for data are not clearly articulated but rather scattered through all of the COBP.

The revised COBP acknowledges the intrinsic value of data by providing Data treatment in its own chapter. Furthermore, the concept of meta data, i.e. “information about information,” is introduced. Additionally, data domains, data sources, and data classes are defined. The overall objective is to establish a new view of data as a strategically valuable entity in its own right. Operational requirements and technical constraints are formulated to enable the establishment of a common data infrastructure thereby providing for the long-term reemployment of data once captured.

However, the revised COBP is still focussed on the domain of conducting a single operational analyses (OA) study. The overarching objective of this paper is to allow the reader to realise the full spectrum of the potential benefits of data standardisation, aligned data engineering processes for the broadening OA community, and the long term goal of an established common data infrastructure, the scope must be broadened beyond the limits of a single study.

A commonly agreed upon data infrastructure does not exist today thereby limiting the utility of data across a wide range of multi-disciplinary studies. The technical objective of this paper is to propose some techniques for managing data in the near term that will allow for the transition to a common methodology of data management resulting in data utility across multiple studies in the future. As more and more data becomes available in open sources, standards must be formulated that will allow for that data to be found, manipulated, used, and stored efficiently. Application of these standards will require a new role in the study team, that of the data engineer, who is not only responsible for the already well known data collection process, but also for the harmonization of all efforts connected to the data, including the evaluation of existing data and meta data as well as updating the meta data for use both within the study and ensuring it is available in usable format for future studies.

To summarise the objectives, this paper focuses on the requirement for and proposes processes of data management at the macro as well as at the study level, which will allow for the future re-use of the data across multi-disciplinary study efforts. To this end, the importance of meta data modelling, the role of the data engineer and the methodologies to be established for a future common data infrastructure will be described in more detail than it is in the revised COBP.

To reach these objectives, the following topics will be discussed:

- Section two provides a practical example highlighting the role of data within an OA study that will be used to demonstrate the necessity to cope with the overarching issue of this paper.
- Section three provides the documentation requirements for data consistency and data traceability within and beyond a single study and the necessity to support data reuse by application of appropriate meta data standards are shown.
- Section four explores the new role of the data engineer on the study team.

- Section five introduces technical constraints and applicable technologies to establish the proposed common data infrastructure.
- Section six summarizes the observations and provides some recommendations for near term implementation that will complement the new data section in the revised COBP.

2.0 A PRACTICAL EXPERIENCE ON THE ROLE OF DATA WITHIN A STUDY

This section depicts some insights and lessons learned from participation within an ongoing NATO feasibility study.

2.1 The NATO Active Layered Theatre Ballistic Missile Defence Feasibility Study

A feasibility study is a critical step in the NATO Phased Armaments Procurement System (PAPS). Essential to the transformation of a NATO Staff Target to a NATO Staff Requirement, it must provide a detailed architecture design and operational performance standard for the project definition phase. The operational analysis conducted in such a study has to be documented thoroughly. Recent national and NATO studies and study results have to be taken into account and should be reused wherever possible. Decisions and associated analyses supporting those decisions have to be documented in a traceable form and should be reusable in follow-on steps of the NATO PAPS.

The example case used here is the ongoing NATO Feasibility Study on Active Layered Theatre Ballistic Missiles Defence (ALTBMD) being conducted on behalf of the NATO Consultation, Command and Control Agency (NC3A). NATO is funding two contracts for the NATO ALTBMDFeasibility Study and the NC3A has invited two consortia of international companies to conduct the feasibility study in parallel. The consortium, from which the examples used in this section have been drawn, combines leading US and European studies and systems houses committed to develop a viable long-term TMD program for NATO: SAIC (US), Boeing (US), Diehl (GE), EADS (FR), IABG (GE), QuinetiQ (UK), and TNO (NL).

Many aspects of the revised COBP are reflected in the ALTBMDFeasibility study. For example, the list of deliverables can be mapped quite easily to the products of an OA study as defined in the revised COBP. Also the methods described in the study dynamics section can be clearly observed. However, this paper will limit itself to those examples derived from participating in the study group relevant to the data section of the COBP.

The ALTBMDFeasibility Study fits in a logical series of NATO study efforts evaluating the military necessity of theatre ballistic missile defence. In 1993, the NATO Council approved the Conceptual Framework for Extended Air Defence followed in 1999 by the refined NATO Air Defence Committee Policy Paper, which further develops concepts for Extended Integrated Air Defence (EIAD). All of this work was supported by respective OA studies and the related data was used to support the ALTBMDFeasibility study findings.

In addition to the NATO studies, a number of national studies have dealt with related issues. For example, the US Ballistic Missile Defence Organisation (BMDO) is a source for a number of significant analyses that have been previously accomplished. Further, in Europe a lot of work has been done, e.g. within the French-Italian SAMP/T programme. Additionally, information can be found in a number of the weapon system programmes themselves, among others the Theatre High Altitude Area Defence (THAAD) programme, the Medium Extended Air Defence System (MEADS) programme and the respective PATRIOT programmes. These limited examples highlight how the efficiencies gained from re-using data from existing sources can provide a rich base for a study effort.

Building up a Common Data Infrastructure

Within the ALTBMD Feasibility Study additional operational analyses are being conducted. These analysis tasks deal with the vulnerability and the survivability of systems, new details in the engagement process of enemy ballistic missiles, the derivation of engagement models for missiles carrying sub-ammunition including nuclear, biological and chemical options, and more ALTBMD related issues. In addition, costs and logistics evaluations are adding their part to the whole study result.

At the end of the efforts, an architecture proposal and inputs for the NATO Staff Requirements will be derived using a variety of different simulation systems and other OA tools – including the TMDSIM, EADSIM and EADTB. Consequently, three requirements have to be fulfilled within the feasibility study:

- The study results of legacy studies from the participating nations and related companies must flow into the actual study design. In addition, the detailed findings of the tasks dealing with vulnerability, ammunition, kill probabilities etc. must eventually find their way into the higher aggregated simulation experiments that will be conducted to evaluate the efficiency of the ALTBMD architectures. Automated tools to convert the data into the needed data formats as well as procedures to assure the data flow would have made the task easier, however, due to the lack of common standards this effort had to be conducted mainly manually.
- As the different tasks of the study all use their own tools and models, the traceability of data is essential. Every data element should be documented, identifying which other study tasks or former studies are related to it and in what form.
- The results of the study – not only in form of a recommended ALTBMD architecture but also all interim steps, detailed results of sub-tasks, evaluated alternatives, etc. – will be reused in the envisaged follow on procurement process. The ability of the data to be effectively reused will depend in large part on how well it is documented in this study and the methods of archiving.

As a result of these requirements, the study team determined that it was necessary to agree on a set of common data standards which would enable the international participants in the study to store and exchange data in a common information repository. The use of the NATO Consultation, Command and Control System Architecture Framework [NATO 2000] helped in structuring the efforts. How this was done can be found in the Simulation Interoperability Standards Organisation (SISO) paper of Adshead, Kreitmair and Tolk [Adshead et al. 2001].

It goes beyond the scope of this paper to detail the solutions used by the NATO ALTBMD Feasibility Study team. However, the role of data within this study can be seen as prototypical for extensive OA study embedded into a greater context of recent, parallel and future studies. The lessons learned from this experience will be summarised in the next subsection.

2.2 Lessons Learned supporting a Common Data Infrastructure

The experiences from the ALTBMD study as well as other similar studies demonstrate the necessity of common standards to support the processes of obtaining, tracing, documenting the changes to, transforming or processing data. These common standards inextricably lead to the need for a special tool that will facilitate these data handling requirements and when implemented will result in reusability of the initial study results in follow-on phases of the current study and for future study efforts.

While the study management team collected and delivered a data package at the beginning of the ALTBMD Feasibility Study that was more complete than previous studies, it nonetheless comprised only a fraction of the data required for the execution of the study. The additional data required had to be obtained by extensive research including mining of the Internet, reading through available recent studies, analysing the input data for the simulation systems and tools that had been used before, etc. Data not only had to be found, it also had to be harmonised within the study team. All these efforts were mainly based on the engineering judgement of subject matter experts (SME's).

Each task group then had to transform the data into the input data needed for the application of tools and models to be used. After the tools and models had processed the data, the results had to be presented to the study team and subsequently had to be delivered to other task groups who needed the results as input parameters (data) for their respective tools and models. Since no common data repository existed, the technical challenge of the required data format transformations and aggregation was exacerbated by the necessity to establish efficient procedures to insure data consistency between the different task groups. To be able to do this, data traceability from the sources through the transformation and aggregation processes had to be assured.

The applicability of the study results and the reusability of the respective data also had to be assured. In the feasibility study this was especially challenging since the transformation of the data from OA study results to operationally usable study data as well as retaining it for later use within the procurement process for consultation, command and control systems had to be assured as well.

As no universally accepted standards were available to support these efforts, a significant effort went into the evaluation and definition of study specific processes to assure that the needed results were obtained. However, even if these developed solutions do become a de facto standard for future NATO ALTBMD studies, a common data infrastructure accompanied by robust technical support will be required to facilitate the execution of the feasibility study significantly. Additional harmonisation will also be required to insure the transparency and usability of the OA study findings in the procurement phases.

The following sections will show what additional efforts can be undertaken to facilitate such data requirements, especially in the context of embedded studies.

3.0 DOCUMENTING DATA USING META DATA

As is demonstrated in the example above and as discussed in the revised COBP data section, after the data requirements are defined three phases for its use within a study can be identified

- Data must be obtained
- Data is used
- Data is delivered

Figure 1 shows the data flow within as well as beyond an OA study including seven steps that will be defined within the descriptions of the three phases.

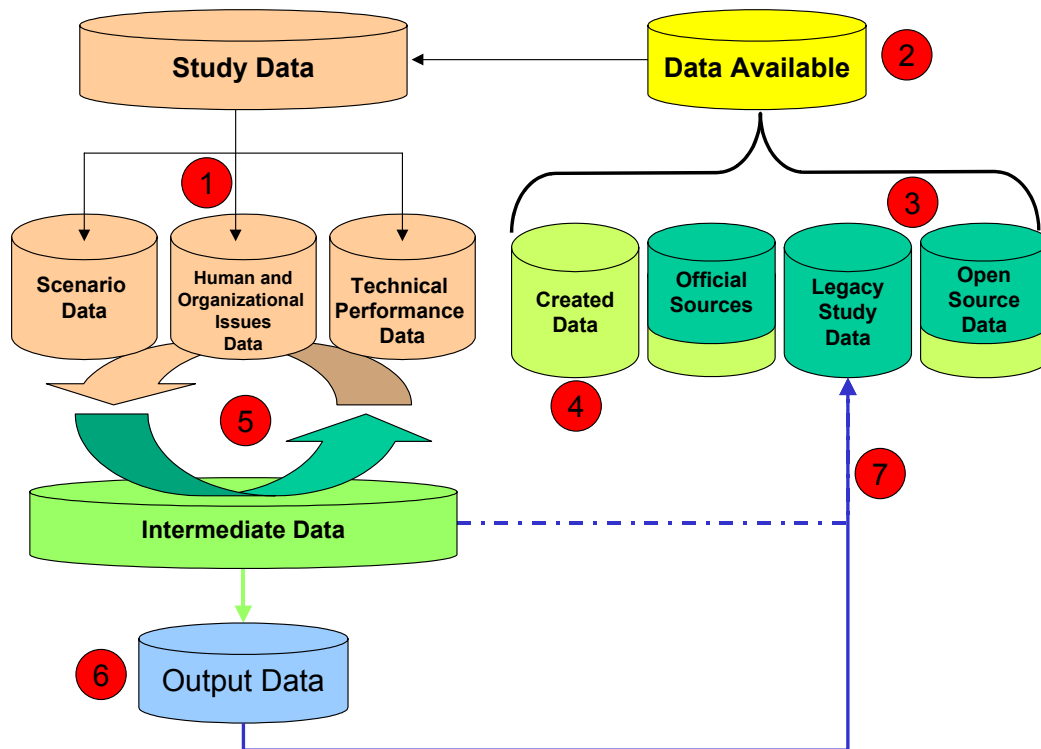


Figure 1: Data Flow within and Beyond an OA Study.

3.1 Obtaining Data

The revised COBP defines four categories of data sources.

- Official Sources are sources such as military databases, other governmental data, data owned by the United Nations, etc.
- Open Sources are data sources that are neither influenced nor controlled by the customer, such as commercial producers (e.g. Jane's) and the Internet.
- Legacy Study Results are data sources derived from other studies conducted by the OA/OR community.
- Finally, when no other means to get the necessary data is available due to the nature of the data requirement or other study constraints data may be estimated by Subject Matter Experts.

Already at each step of the obtaining process, data must be documented to ensure the traceability of results, communicate any constraints connected to the data, and describe any special concerns or requirements for validity, etc. For each data element, the source has to be included in the meta data. If the meta data is not available for the source itself, it should be derived as accurately as possible for each data element or coherent group of data elements. At a minimum the source, reliability of the source, constraints such as models and tools used for processing, title of study, reference to the Internet page should be documented.

To summarise, within this phase, the data have to be defined first (step 1), then the available data has to be checked for consistency and completeness (step 2). Using the various data sources, the data package needed for the study is prepared (step 3), including estimation of not otherwise obtainable data (step 4).

3.2 Data Use

The use of data within the study can be divided into sub-steps that can be of fractal structure within the study itself. First, generally the data obtained must be transformed and aggregated to be useful as input data for a tool or model to be applied in the context of the study. The transformation and aggregation processes of the input data must be documented. As a minimum, the traceability from the obtained data to the input data has to be assured by the meta data documentation allowing the study team to re-evaluate all results connected to input data that is changed during the conduct of the study.¹

By applying tools and models, new data is produced. For these data elements, the tool or model used to provide them as well as the data being used to drive the tool or the model have to be captured in the accompanying meta data. It is not sufficient just to track the tool or model used, even if it is a previously verified, validated and accredited model, since the input data is important for the validity and reliability of the results as well. This must be accomplished for the entire system for each use.

In figure 1, these processes are covered by step 6: data use and transformation within the study.

3.3 Data Delivery

When the input and intermediate data is finally transformed into data supporting the delivered study result the underlying assumptions, constraints, etc. must be documented. The transformation of input and intermediate data is normally accomplished by interpreting the measure of merits to evaluate the essential elements of analysis (e.g. critical questions, critical operational issues, etc.). In all cases in order to ensure that future analysts are able to evaluate the usability of the study results (data) for their studies the underlying assumptions, constraints, etc. have to be sufficiently documented for them to be able to make value judgements regarding data utility.

The same should also be true for the interim results of a study since it is possible that they may be valuable input parameters for future studies as well, although they may just be a by-product of the ongoing OA effort.

In figure 1, this is covered by step 6 (preparing the data for the study report) and step 7 (preparing intermediate and output data for future re-use).

Finally, it is worth thinking about “sanitised” versions of the study results. In the case of classified studies it would be valuable if unclassified insights that could be valuable inputs for the broader OA community could be collected. The accompanying meta data should then contain the reference to the classified study to assure the accessibility in case of need.

In summary, the use of meta data modelling not only enables efficient data traceability and delivers the needed documentation within an individual study, it is also a requirement for efficient data reusability among different studies. Meta data comprises all information about the data needed to search for and evaluate its applicability for a given study purpose.

4.0 DATA ENGINEERING

Until recently, the concerns about data could generally be limited to developing a data collection plan at the beginning of the study. As the preceding three sections illustrate, data’s importance to both an

¹ E.g., in the ALTBMD the vulnerability of a special missile type changes due to some technical break through in the engagement phase, all simulation results using the old vulnerability model (including former studies) have to be at least re-evaluated. In some cases it may even be possible that old study results are not valid any longer.

Building up a Common Data Infrastructure

individual study and to the body of corporate knowledge is increasing daily. Consequently a new sense of professionalism has to be adopted by the OA community concerning the handling of data. The definition of a new role within the OA community as a whole and in the study team in particular is the logical consequence – the data engineer.

The data engineer is responsible for the overall management of data within the context of an individual study and for ensuring that it is properly collected, tagged and archived for later use. Within a specific study effort, the data engineer is responsible for obtaining the data, evaluating the meta data with concern to the study needs, transforming it to meet the tool and model requirements, documenting the data as it is transformed throughout the study effort, conducting meta data modelling to handle the meta data for the study as well as for future studies and for the data and information exchange between the study team and the OA community.

A data engineer is obviously much more than a data collector, although this is still an important task for him. The data engineer must be able, however, to “dig for the data” within the full spectrum of available sources. To effectively do so, this person must not only understand the data itself, but he also must be aware of the macro level data needs of the study. Among other things the engineer must be able to identify the needed level of reliability, acceptable sources, needed formats, fidelity requirements, possibilities for aggregations and deaggregations, limits of data transformation, etc. The data engineer must be able to understand and analyse information repositories of other research communities as well as using the principles of Information Resources Dictionary Systems (IRDS) to map the available data to his own needs.

The data engineer can be seen as the bridge between the OA study team and the data available. The engineer’s job is to assist the study team in finding and obtaining needed data “wherever and in whatever format it should be” to enable them to conduct the study. The data engineer might be compared to the expert within the response cell (RC) of a computer-assisted exercise (CAX) – he must understand the needs and plans of the study team as the RC expert must understand the needs and procedures of the training audience. The data engineer must also know where and how to obtain the data and transform it to the needs of the study team just as the RC expert has to generate the appropriate simulation system inputs from the commands of the training audience.

The data engineer will be supported by new data management tools like improved search engines, meta crawlers, etc. analogous to the way software support, like automatic interfaces between the simulation system and the command and control system, facilitates the work of the RC expert.

5.0 THE COMMON DATA INFRASTRUCTURE

As pointed out before, one of the main problems the broadening OA community is faced with is the heterogeneity of data sources being used. This is not a new problem. The necessity to agree on common standards is one of the driving factors for the Simulation Interoperability Standards Organisation (SISO). Similar recommendations can also be found within the Military Operations Research Society (MORS). The following citation is taken from the conclusions of the MORS Data Working Group, and although it is over ten years old it is still valid:

“The single most important activity ... would be a concerted effort to get all members of the team to see the same battlefield through a common engineering approach, shared data-bases, common tool sets, and a network of all players. It was consensus of the working group that one of the most critical needs was to produce an overt structure that linked all members of the data/modelling team. ... The data sets must be clearly described and understandable to a user with subject matter knowledge ... The data description must be robust enough to inspire user confidence in the data.” [DWG 1988]

As pointed out in the COBP and in previous sections of this paper, the overarching objective regarding data is the seamless sharing of information between:

- the study team members
- the evolving phases of the study
- the models and tools used within the study
- the study team and the broader OA community (reusability).

Documentation of data (including validity and reliability of sources, constraints, etc.), consistent recording of data transformation and enabling data re-use of both the interim and final study findings by future studies are the imperatives behind the drive to establish a common data infrastructure. The technical feasibility of such a common infrastructure has already been proven in the domain of electronic commerce. The obvious similarity between the applications of Collaborative Product Commerce and the Support of Combined and Joint Military Operations Other Than War has been shown (e.g. Krusche and Tolk 2000). The necessary technologies are based on the idea of efficient shared data management using the same procedures and meta data models to document the findings of these processes. The common data infrastructure has to be able to store the data as well as the meta data in a well defined – and preferably standardised – manner. Fortunately, a mature international standard is already established that can be applied to serve the OA community's need – an Information Resource Dictionary System (IRDS). The main ideas of an IRDS are defined in the ISO IRDS standard [ISO 1990]. The main purpose of an IRDS is to support data administration and data management. A NATO application example can be found in [NDAG 1999]. Another existing source of collected data is the US Defence Modelling and Simulation Office's (DMSO) Authoritative Data Source (ADS) Project. The ADS project catalogues all M&S relevant data/knowledge sources within the US Department of Defence and the Modelling and Simulation community at large.

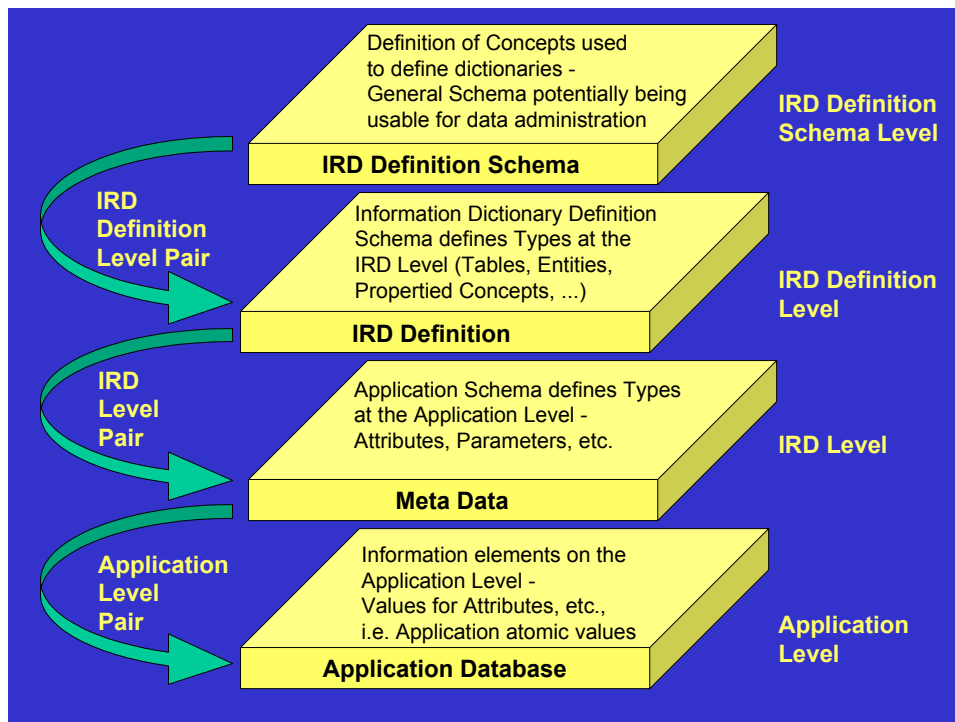


Figure 2: Levels of Information in IRDS.

Building up a Common Data Infrastructure

An IRDS can be defined as a software system comprising and managing the information resource dictionary in which the information of all participating applications will be recorded. It has been shown how this idea can be extended in the way that the IRDS can also be used to support the federate integration process of the high level architecture (HLA) by making the efforts of the data standardisation community usable for the federation builders.

The IRDS framework defines four levels of information shown in figure 2. Each level in the framework has a sub-level that consists of the definition of the information contained in its respective sub-levels. Therefore, the use of the ISO IRDS framework allows a gradual introduction of concepts and methodologies from the most abstract form down to most concrete and tangible application and implementation requirements. Thus, the different methodologies of relational data modelling using IDEF1X, and object oriented modelling using UML are nothing more or less than different concepts within the IRDS on the respective level. By storing the respective data management results also within the IRDS, the IRDS builds the kernel for a common data infrastructure fulfilling the needs as stated before. If the needed data is available in whatever format using whatever data modelling methodology, it can be found and transformed in standardised manner from the IRDS respective the common data infrastructure.

In addition to these technologic solutions, data management is necessary. Within NATO, data management is defined as planning, organising and managing of data by defining and using rules, methods, tools and respective resources to identify, clarify, define and standardise the meaning of data as of their relations. This results in validated standard data elements and relations, which are going to be represented and distributed as a common shared data model. As this definition indicates and as this paper and the revised COBP support, efficient data administration is an information intensive process involving a wide range of participants with impact and implications that extend well beyond the scope of a single study. The data required is generated, managed, and used by a large number of participants in the multi-disciplinary and multi-national study team as well as by members of the broader OA community. Every entity delivering an application to participate in multiple federations – consuming and delivering data from and for the federation – has to be involved in the process of data management. Effective collaboration between all participants in the process of establishing a common data standardisation is essential in order to gain and preserve a common understanding of shared data. Therefore, an essential purpose of data administration activities must be to achieve an integrated data standard that will facilitate the broader needs of the OA community for data use/reuse.

It should be pointed out that the requirements for aligning the data management procedures of the OA community – and in many cases even to make the necessity of data management and documentation clear to the decision makers – are at least as challenging as the technical ones. However, the benefit for the OA community is expected to be very high.

6.0 CONCLUSIONS AND RECOMMENDATIONS

The Data Section within the revised COBP has been a valuable addition to the first version. It will help to make the analysts, users and the decision makers aware of the strategic value assigned to re-usable and shared data. The necessity for a common data infrastructure – accompanying other repositories like a model and tools repository as recommended in the NATO Long Term Scientific Study on Human Behaviour Representation [NATO 2001] – is becoming obvious.

As the OA community is broadened to take into account human and organisational issues in addition to technical performance as part of the equation to evaluate the military socio-technical system, the existing common basis of OA and modelling and simulation must likewise be broadened to include the research domains of psychology, sociology and other human sciences. It is essential to co-ordinate standardisation efforts as early as possible to avoid repetitive work and to enable information sharing across the broadened OA Community.

A common data infrastructure using a standardised way to use, modify and record data elements is a necessary requirement for efficient and continuously interoperable information sharing within the broad OA community. Success in establishing such a data infrastructure through the application of the techniques outlined in the revised COBP for current and future studies will contribute greatly to assuring the success of future joint and combined efforts across the full spectrum of military operations.

7.0 REFERENCES

Following reports and articles are referenced in this paper:

[Adshead et al. 2001] S. Adshead, T. Kreitmair, A. Tolk: “Definition of ALTBMD Architectures by Applying the C4ISR Architecture Framework”, Paper 01F-SIW-112, Proceedings of the Simulation Interoperability Workshop Fall 2001, Orlando, Florida, September 2001.

[DWG 1988] Report of the Data Working Group, Simulation Technology-1997 (SIMTECH-’97), US Army Engineer Studies Centre, March 1988; Monterey, California, June 1988; National Defence University, November 1988. (DTIC: ADB 152 051 and ADB 152 052 ADB 152 053).

[ISO 1990] ISO Standard IS10027: IS10027:1990, “An Information Resource Dictionary System (IRDS) Framework”, 1990.

[Kendrick et al. 1999] David E. Kendrick, Jack Sheehan, Lana Eubanks McGlynn, Mike Hopkins: *Authoritative Data Sources for use in DoD Modelling and Simulation*, Proceedings of the Simulation Interoperability Workshop Fall 2001, Orlando, Florida, Fall 1999.

[Kendrick et al. 2000] David E. Kendrick, Jack Sheehan, Mike Hopkins, Katherine Rowe: *Authoritative Data Sources: Enhancements to Satisfy Future Needs and Long Term Maintenance Requirements*, Proceedings of the Simulation Interoperability Workshop Fall 2001, Orlando, Florida, Spring 2000.

[Krusche and Tolk 2000] Stefan Krusche, Andreas Tolk: “Information Processing as a Key Factor for Modern Federations of Combat Information Systems”, NATO/RTO Information Systems Technology Panel (IST) Symposium “New Information Processing Techniques for Military Systems”, Istanbul, October 2000.

[NATO 2000] NATO Consultation, Command and Control Board (NC3B): “NATO C3 System Architecture Framework”, AC/322-WP/0125, Brussels, November 2000.

[NDAG 1999] NATO Consultation, Command and Control Board (NC3B), Information Systems Sub-Committee (ISSC), NATO Data Administration Group (NDAG): “Feasibility Study for the NATO C3 Information Resource Dictionary System (NC3 IRDS)”, AC/322(SC/5-WG/3)WP4, Brussels, September 1999.

8.0 LIST OF ACRONYMS

Following acronyms and abbreviations are used within this paper:

ALTBMD	Active Layered Theatre Ballistic Missile Defence
C3	Consultation, Command and Control
COBP	Code of Best Practise
EADSIM	Extended Air Defence Simulation
EADTB	Extended Air Defence Testbed

Building up a Common Data Infrastructure

EEA	Essential Elements of Analysis
EIAD	Extended Integrated Air Defence
HLA	High Level Architecture
ICAM	Integrated Computer-Aided Manufacturing
IDEF1X	ICAM Definition for Data Modelling
IRDS	Information Resource Dictionary System
MEADS	Medium Extended Air Defence System
MOE	Measure of Effectiveness
NC3A	NATO C3 Agency
NC3B	NATO C3 Board
NDAG	NATO Data Administration Group
NSR	NATO Staff Requirement
NST	NATO Staff Target
OA	Operational Analysis
PAPS	Phased Armaments Procurement System
SAMP/T	Sol-Air Moyenne-Portée/Terrestre
SISO	Simulation Interoperability Standards Organisation
SIW	Simulation Interoperability Workshop
THAAD	Theatre High Altitude Area Defence
TMDSIM	Tactical Missile Defence Simulator
UML	Unified Modelling Language

AUTHOR BIOGRAPHIES

Mark R. Sinclair is the Advanced Programs Manager for Veridian's Simulation and Integrated Solutions Group. He participated as a member of the US delegation to the NATO SAS-026 Panel charged with revising and extending the *NATO Code of Best Practice for Command and Control Assessment*. He is a retired Lieutenant Colonel in the United States Marine Corps where he served as a Command and Control Systems officer for more than twenty-five years. His operational experience includes Vietnam, Central America, Haiti and the Balkans.

Andreas Tolk is Senior Scientist at the Virginia Modeling Analysis and Simulation Center (VMASC) of the Old Dominion University, Norfolk, VA, since April 2002. Before, he was Vice President for Land Weapon Systems at IABG mbH in Ottobrunn, Germany. He participated as a member of the German delegation to the NATO SAS-026 Panel. Within the referenced NATO Feasibility Study, he was responsible for Command and Control Architecture Definitions. He is an expert for Command and Control Systems and Simulation Systems Interoperability issues.

Building up a Common Data Infrastructure

Mark R. Sinclair
Veridian

Dr. Andreas Tolk
**Virginia Modeling, Analysis
& Simulation Center (VMASC)**



- Introduction
- A Practical Experience on the Role of Data within a Study
- Documenting Data using Meta Data
- Data Engineering
- The Common Data Infrastructure
- Conclusions and Recommendations

Introduction

- Role of Data
- Value of Data
- Data Management

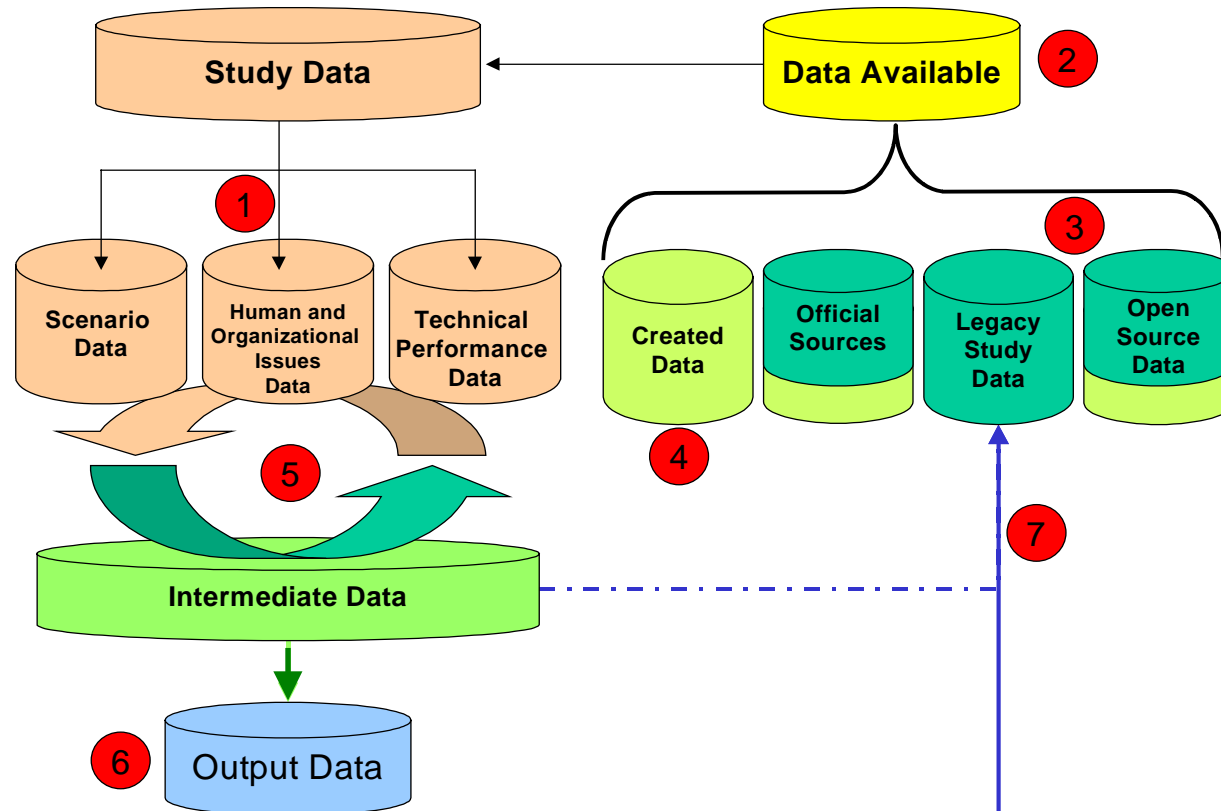
A Practical Experience on the Role of Data Within a Study

- Case study – NATO Active Layered Theatre Ballistic Missile Defence Feasibility Study
 - Nature of the Study
 - Multinational
 - Based on great number of recent studies
 - Study Data Requirements
 - Use of legacy data (NATO and nations legacy studies)
 - Data traceability (connection of input and result)
 - Data reusability (for next step in PAPS)
- Lessons Learned
 - Urgent Need for Common Data Standards
 - Need for a Common Data Infrastructure

Documenting Data Using Meta Data

Data (Re-)Use

Data Acquisition

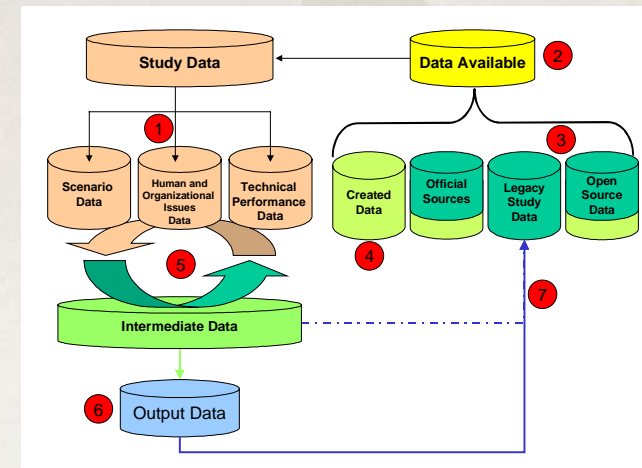


Data Transformation

Data Delivery

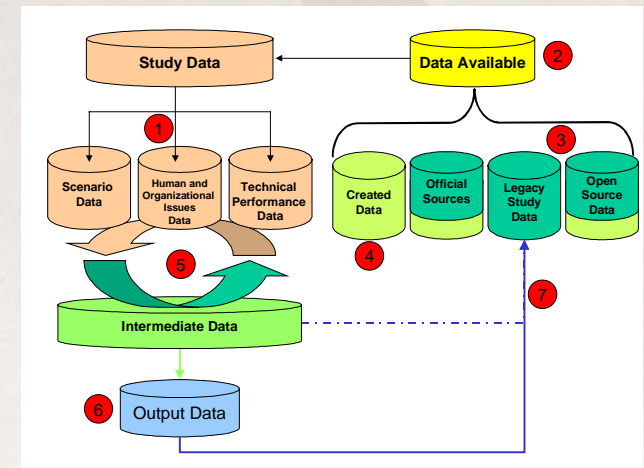
Data Acquisition

- Official Sources
- Open Sources
- Legacy Study Results
- Created Data (SME)
- Documentation
- Included in this Phase - Steps 1, 2, 3,& 4.



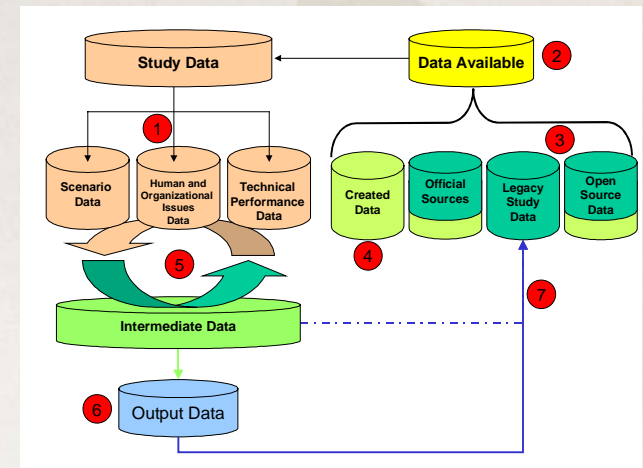
Data (Re-)Use and Transformation

- Transformation & Aggregation
 - Documentation
- New Data Production
 - Model & Tool Application
 - Documentation
- Included in this Phase – Steps 5 & 6.



Data Delivery

- Final Study Data Accumulated
- Intermediate Data Identified
- Assumptions, Constraints, etc. Documented
- Sanitized versions
- Included in this Phase - Steps 5,6,& 7.

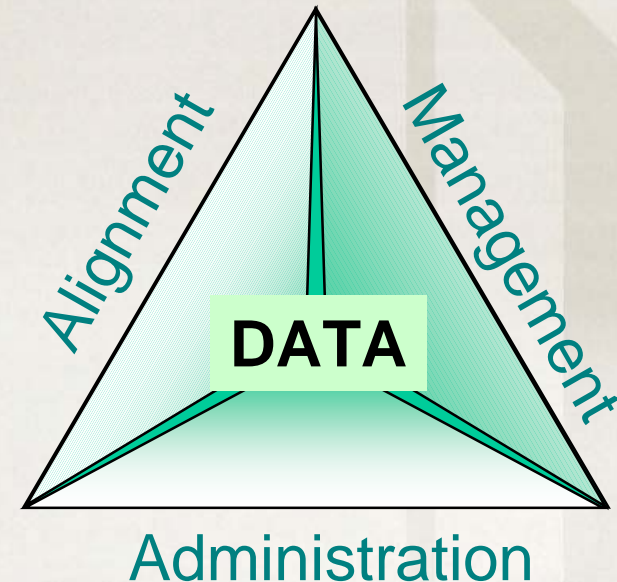


Data Engineering

➤ Data Engineering

- Overall Data Management
- Collection
- Tracing
- Documentation
- Validation (VV&C)
- Research
- Archivist

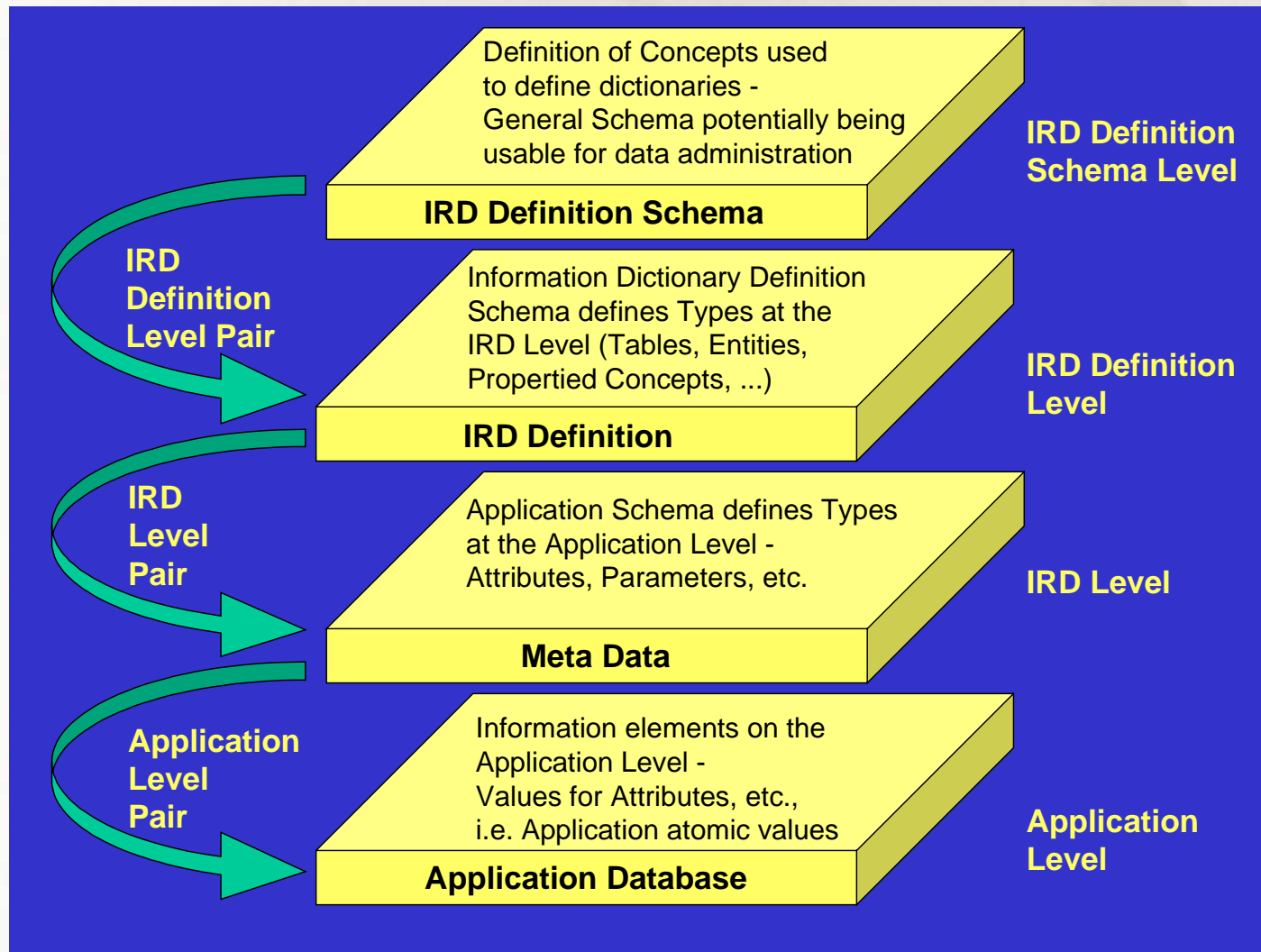
➤ Responsible for BOTH current study and future utility of the data.



Common Data Infrastructure

- Common Standards
- Seamless sharing
- Documentation
- Information Resource Dictionary System (IRDS)
- *Authoritative Data Source (ADS) Project*
- *Data Management*

Excursus: Information Resource Dictionary System



ISO Standard IS10027:1990

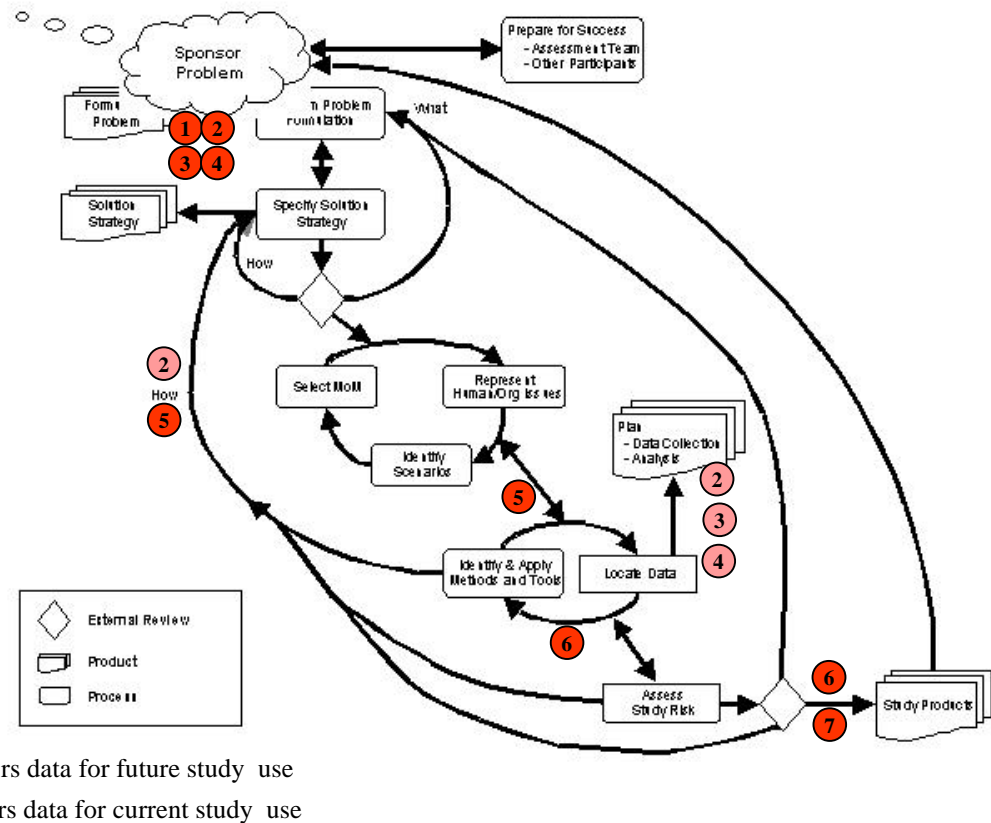
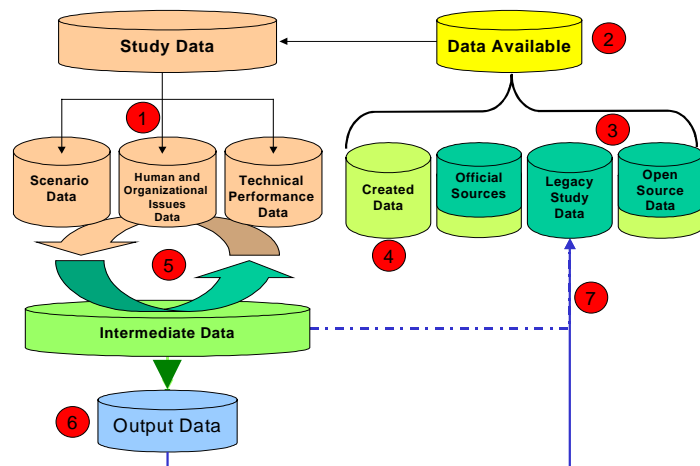
Common Data Infrastructure

- Common Standards
- Seamless sharing
- Documentation
- Information Resource Dictionary System (IRDS)
- Authoritative Data Source (ADS) Project
- Data Management

Conclusions and Recommendations

- Data's value is greater than its utility for a single study
- Data is reusable
- Common Data Infrastructure is necessary
- Broader Research Domains required to support military OA
- Application of the Revised COBP enhances the utility of data across the broad community of interest.

Data and the SAS-26 Study Process



Questions

